

METHODS AND SYSTEMS FOR USING XML SCHEMAS TO IDENTIFY AND CATEGORIZE DOCUMENTS

Abstract of the Disclosure

5 A method for identifying an XML document includes the steps of obtaining the document, matching the document against a plurality of XML schemas that specify a set of document types that support a particular application, and, based on the results of these comparisons, outputting information regarding the document type. The outputted information could include information regarding the identity of the document type.

10 Furthermore, in the event that the document fails to match the schemas exactly, the document type which most closely matches the given document could be identified. In this case, a match score for the closest document might also be returned. A match score of zero could indicate a perfect match and any positive value a mismatch, with the score value increasing with the degree of mismatch, for example.

15

20

25

30